

Appendices

Datatypes

This section contains the datatype definitions used within the Public Health Conceptual Data Model. The datatypes are drawn from the HL7 Reference Information Model, and represent a subset of the datatypes defined therein.

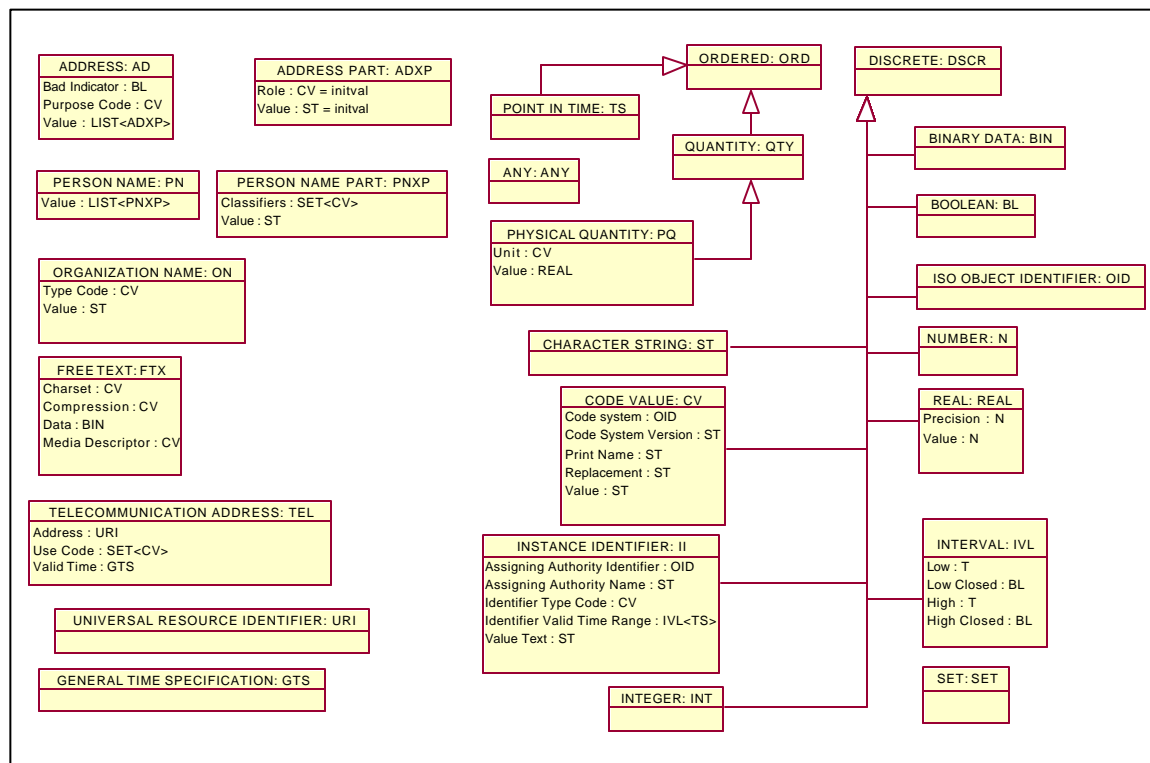


Figure 1. Datatypes Diagram

The classes and attributes that comprise the datatypes are described below.

Datatype: ADDRESS : AD

Is a Composite Datatype

Description of: AD

This Address datatype is used to communicate postal addresses and residential addresses. The main use of such data is to allow printing mail labels (postal address), or to allow a person to physically visit that address (residential address). The difference between postal and residential address is whether or not there is just a post office box. The residential address is not supposed to contain other

information that might be useful for finding geographic locations or doing epidemiological studies. These addresses are thus not very well suited for describing the locations of mobile visits or the "residency" of homeless people.

Components of: **AD**

Bad Indicator : BL

Indicates that this address is not working

Purpose Code : CV

A purpose code indicates the use for a given address. Examples might include preferred residency (used primarily for visiting), temporary (visit or mailing, but see History), preferred mailing address (used specifically for mailing), and some more specific ones, such as "birth address" (to track addresses of small children). An address without specific purpose code might be a default address useful for any purpose, but an address with a specific purpose code would be preferred for that respective purpose.

Value : LIST<ADXP>

This contains the actual address data as a list of address parts that may or may not have semantic tags.

Datatype: **ADDRESS PART : ADXP**

Is a Composite Datatype

Description of: **ADXP**

This type is not used outside of the Address datatype. Addresses are regarded as a token list. Tokens usually are character strings but may have a tag that signifies the role of the token. Typical parts that exist in about every address are ZIP code, city, country but other roles may be defined regionally, nationally, or on an enterprise level (e.g., in military addresses). Addresses are usually broken up into lines that are indicated by special line break tokens.

Components of: **ADXP**

Role : CV

The role of an address part (if any) indicates whether an address part is the ZIP code, city, country, or post office box.

Value : ST

The value of an address part includes the text for the specific component of the address. It is what is printed on an address label.

Datatype: ANY : ANY

Is a Primitive Datatype

Description of: ANY

This is a generalized datatype that represents any other datatype within the model. This concept is needed to support observation values, and to let those values take on any datatype.

Datatype: BINARY DATA : BIN

Is a Primitive Datatype

Has Super Types: **DSCR**

Description of: BIN

Binary data is a sequence of uninterpreted raw bytes (8 bit sequences, or octets).

Datatype: BOOLEAN : BL

Is a Primitive Datatype

Has Super Types: **DSCR**

Description of: BL

The boolean type stands for the values of two-valued logic. A boolean value can be either true or false.

Datatype: CHARACTER STRING : ST

Is a Primitive Datatype

Has Super Types: **DSCR**

Description of: ST

A string of characters where every character is represented by a uniquely identifiable entity within the string.

Datatype: **CODE VALUE : CV**

Is a Composite Datatype

Has Super Types: **DSCR**

Description of: **CV**

A code value is exactly one symbol in a coding system. The meaning of the symbol is defined exclusively and completely by the coding system from which the symbol originates.

Components of: **CV**

Code system : OID

An object identifier referring to the code system that defines the code value. The OID supports unambiguous reference to standard coding systems - including HL7 codes, as well as to local codes.

Code System Version : ST

A version descriptor defined specifically for the given coding system.

Print Name : ST

A sensible name for the code as a courtesy to an interpreter of the message. The name should not be considered as carrying the meaning of the code, it should never be sent alone, and it does not modify the meaning of the code.

Replacement : ST

A name for the concept whose meaning is being conveyed. The replacement is used if the concept cannot be captured by a code in the specified coding system. If the value attribute is set, the replacement attribute **MUST NOT** be set. In no way can a replacement string modify the meaning of the code value.

Value : ST

This is the plain symbol. E.g., "784.0"

Datatype: **DISCRETE : DSCR**

Is a Primitive Datatype

Has Sub Types: **ST**
REAL
OID

N
IVL
INT
II
CV
BL
BIN

Description of: **DSCR**

Abstract generalized type for any discrete type.

Datatype: **FREE TEXT : FTX**

Is a Composite Datatype

Description of: **FTX**

This free text datatype can convey any data whose primary purpose is to be shown to people for interpretation. Free text can be any kind of text, whether written language (formatted or unformatted) or multi-media data.

Components of: **FTX**

Charset : CV

Definition of the character encoding if different from the default encoding.

Compression : CV

Indicates that the raw byte data is compressed, and which compression algorithm is being used.

Data : BIN

Contains the free text data as raw bytes.

Media Descriptor : CV

Allows selection of the appropriate free text data. The default value is "text/plain".

Datatype: **GENERAL TIME SPECIFICATION : GTS**

Is a Primitive Datatype

Description of: **GTS**

This is a primitive datatype that is conceptually an arbitrary set of points in time. It is any combination of 1) a point in time, and 2) an interval of time. This includes

uncertain points and intervals of time. The contents of a GTS instance contains values that are defined in terms of a literal expression syntax that allows statement of any needed frequency or time pattern.

For example, these are some ways in which the GTS datatype is used within the PHCDM:

Dates are represented based on the precision needed and/or supplied. Y1999 represents the year 1999, while Y199909 indicates September, 1999, and Y19990926 indicates September 26, 1999. If only month and day are available, M0926 indicates September 26. The following are some of the period identifiers available: Y for year, M (or MY) for month of the year, D (or DM) for day of the month, H (or HD) for hour of the day.

The tag WY (w for week and y for year) indicates week of the year. Therefore WY23 is the 23rd week of the year.

Date ranges are indicated by a pair of dates separated by a dash, "-". For example, "Y20000110 - Y20000204" indicates a period beginning on January 10, 2000 and ending on February 4, 2000. Open-ended periods can be indicated by a date either preceded by a dash (period ending on the date indicated), or by a date followed by a dash (period beginning on the date indicated).

Durations are indicated by square brackets, "[" and "]". The duration amount and unit are within the brackets. For example "[10 min]" indicates 10 minutes, and "[3 day]" indicates 3 days.

These expressions can be concatenated together to express the union of multiple time concepts. For example, "M09 D26" is an alternate way of indicating September 26. Also, "Y20000224 [8 hour]" indicates a duration of 8 hours on February 24, 2000.

As a general statement, the construction of GTS instances is based on the following ideas:

- Singular time intervals as continuous sets of time points, specified through low and high boundary or width (in case no boundary is known.)
- Periodic time intervals as discontinuous sets of time points, specified through a period duration and a time offset (phase) interval.
- The set-operations intersection and union on such continuous and discontinuous sets of time to form arbitrary sets of time.
- The reduction of any arbitrary set of time into an outer bound interval and a sequence of occurrence intervals, no matter how complex the definition of this arbitrary set is.

- The use of probability distribution datatypes to account for the uncertainty in scheduling and time orders, or, in other words, to allow "fuzzy" constraining of time sets.

Time can be specified in terms of an absolute even flow of time, or events taking place in time can be aligned to calendars.

Datatype: INSTANCE IDENTIFIER : II

Is a Composite Datatype

Has Super Types: **DSCR**

Description of: II

The datatype is used to uniquely identify an entity that exists within a computer system or other well-controlled identification scheme.

Components of: II

Assigning Authority Identifier : OID

The ISO object identifier for the organization or identifier issuing scheme that is responsible for the integrity and validity of the identifier. This field guarantees the uniqueness of the identifier, and permits the origin of the identifier to be determined. If the organization uses OIDs for internal object identifiers, this may be the only field valued.

Assigning Authority Name : ST

The name of the organization or scheme responsible for the identifier.

Identifier Type Code : CV

A code representing the type of identifier. For example, the code might represent the US national provider ID, US national payer ID, medical record number, and social security number.

Identifier Valid Time Range : IVL<TS>

The time range during which the identifier is valid. It may be undefined on either side since in some cases only the start date for ID validity will be known, while, in others, only the end date is available.

Value Text : ST

The character string value of the identifier. For example the character string "123-45-6789" for a medical record number.

Datatype: **INTEGER : INT**

Is a Primitive Datatype

Has Super Types: **DSCR**

Description of: **INT**

Integer numbers are precise numbers that are results of counting and enumerating. The set of integers is infinite but countable. Two special integer values are defined for positive and negative infinity.

Datatype: **INTERVAL : IVL**

Is a Generic Datatype

Has Super Types: **DSCR**

Description of: **IVL**

Generic datatype that can express a range or interval of values. An interval is a set of consecutive values of any totally ordered datatype. An interval is thus a continuous subset of its base datatype.

Datatype: **ISO OBJECT IDENTIFIER : OID**

Is a Primitive Datatype

Has Super Types: **DSCR**

Description of: **OID**

The ISO Object Identifier is defined by ISO.

Datatype: **NUMBER : N**

Is a Primitive Datatype

Has Super Types: **DSCR**

Description of: **N**

The representation of a number. It is used as a generalized type for different numeric representations.

Datatype: **ORDERED : ORD**

Is a Primitive Datatype

Has Sub Types: **TS**
 QTY

Description of: **ORD**

Abstract generalized type that at least contains naturally ordered subsets.

Datatype: **ORGANIZATION NAME : ON**

Is a Composite Datatype

Description of: **ON**

A name for an organization, such as "Centers for Disease Control and Prevention".

Components of: **ON**

Type Code : CV

A code identifying the use for an organization name. Possible values include: L - Legal, A - Alias, D - Display, ST - Stock Exchange.

Value : ST

The actual name data as a simple character string.

Datatype: **PERSON NAME : PN**

Is a Composite Datatype

Description of: **PN**

A Person name is one full name of a person. A name such as "Jim Bob Walton, Jr." is one instance of a Person name. The parts of this name "Jim", "Bob", "Walton", and "Jr." are person name parts.

Components of: **PN**

Value : LIST<PNXP>

This contains the actual name data as a list of name parts that may or may not have semantic tags.

Datatype: **PERSON NAME PART : PNXP**

Is a Composite Datatype

Description of: **PNXP**

This type is not used outside of the Person Name datatype. Person Names are regarded as token lists. Tokens usually are character strings but may have a tag that signifies the role of the token. Typical name parts are given names and family names; other part types may be defined culturally.

Components of: **PNXP**

Classifiers : SET<CV>

Classifications of a name part. One name part can fall into multiple categories, such as given name vs. family name and name of public record vs. nickname.

Value : ST

The value of a name part.

Datatype: **PHYSICAL QUANTITY : PQ**

Is a Composite Datatype

Has Super Types: **QTY**

Description of: **PQ**

A physical quantity results from a measurement act. It consists of a value and a unit.

Components of: **PQ**

Unit : CV

The unit of measure. Typically this is a unit, such as kilograms or miles per hour, that is drawn from a table of units of measure. Note that "count" is also included as a unit. This is used, for example, when collecting information about the number of interventions of a particular type.

Value : REAL

The magnitude of the quantity measured in terms of the unit.

Datatype: **POINT IN TIME : TS**

Is a Primitive Datatype

Has Super Types: **ORD**

Description of: **TS**

A point in time is a scalar defining a point on the axis of natural time.

Datatype: **QUANTITY : QTY**

Is a Primitive Datatype

Has Super Types: **ORD**

Has Sub Types: **PQ**

Description of: **QTY**

Abstract generalized type for any quantitative type.

Datatype: **REAL : REAL**

Is a Composite Datatype

Has Super Types: **DSCR**

Description of: **REAL**

A numerical amount. In order to facilitate computer representation, this is, by assumption, a floating-point number.

Components of: **REAL**

Precision : N

The precision of the floating point number in terms of the number of significant decimal digits.

Value : N

The value, expressed as an integer.

Datatype: **SET : SET**

Is a Generic Datatype

Description of: SET

SET is an unordered collection of unique items.

Datatype: TELECOMMUNICATION ADDRESS : TEL

Is a Composite Datatype

Description of: TEL

This is a token assigned as a mechanism for locating a telecommunication device such as a telephone, website, or email address.

Components of: TEL

Address : URI

This is an arbitrary address string that uniquely identifies an address in a particular domain.

Use Code : SET<CV>

The purpose of the "use code" is to advise in a system or user's selecting an appropriate telecommunication address to reach a party for a given telecommunication need. The following mandatory value domain is defined: PR - primary residence (home) OR - other residence (other home) WP - work/business/office communication address VR - vacation residence AS - automated answering service EC - emergency contact BP - beeper/pager CL - cellular/wireless phone

Valid Time : GTS

This is a General Time Specification (GTS) that identifies the periods of time during which this telecommunication address can be used. For a telephone number this can indicate the time of day in which the party can be reached on that telephone. For a web address, it may specify a time range in which the web content is promised to be available under the given address

Datatype: UNIVERSAL RESOURCE IDENTIFIER : URI

Is a Primitive Datatype

Description of: URI

The URI is used to refer to addresses of communicating entities used in order to transmit any kind of information. This may be used for messaging addresses, and for non-computer communication.